

Decristan, Jasmin; Kunter, Mareike; Fauth, Benjamin; Büttner, Gerhard; Hardy, Ilonca; Hertel, Silke
What role does instructional quality play for elementary school children's science competence? A focus on students at risk

Journal for educational research online 8 (2016) 1, S. 66-89



Empfohlene Zitierung/ Suggested Citation:

Decristan, Jasmin; Kunter, Mareike; Fauth, Benjamin; Büttner, Gerhard; Hardy, Ilonca; Hertel, Silke: What role does instructional quality play for elementary school children's science competence? A focus on students at risk - In: Journal for educational research online 8 (2016) 1, S. 66-89 - URN: urn:nbn:de:0111-pedocs-120321

in Kooperation mit / in cooperation with:

WAXMANN
VERLAG GMBH
Münster · New York · München · Berlin



<http://www.waxmann.com>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Jasmin Decristan, Mareike Kunter, Benjamin Fauth,
Gerhard Büttner, Ilonca Hardy & Silke Hertel

What role does instructional quality play for elementary school children's science competence?

A focus on students at risk

Abstract

Research on classroom instruction has consistently identified characteristics that contribute to student learning. For instance, these include structural-organizational aspects (e.g., classroom management) and affective aspects (e.g., classroom social climate). The idea that the effects of instruction may differentially depend on students' characteristics has been investigated within the scope of aptitude-treatment-interactions (ATI) research. This study of elementary school (1,041 students, 54 classes) builds on ATI and examines main effects and interaction effects of instructional quality (i.e., classroom management and classroom social climate) and individual risks of school failure (i.e., demographic risk: immigration background or functional risk: low cognitive ability scores) on students' science competence. Based on hierarchical linear modeling and class-level

Dr. Jasmin Decristan (corresponding author), German Institute for International Educational Research (DIPF), Department of Educational Quality and Evaluation, Schloßstraße 29, 60486 Frankfurt am Main, Germany & Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt am Main, Germany
e-mail: decristan@dipf.de

Prof. Dr. Mareike Kunter · Prof. Dr. Gerhard Büttner, Faculty of Psychology and Sports Sciences, Goethe-University, Theodor-W.-Adorno-Platz 6, 60629 Frankfurt am Main, Germany & Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt am Main, Germany
e-mail: kunter@paed.psych.uni-frankfurt.de
buettnner@paed.psych.uni-frankfurt.de

Prof. Dr. Benjamin Fauth, Hector Research Institute of Education Sciences and Psychology, University of Tuebingen, Europastraße 6, 72072 Tübingen, Germany
e-mail: benjamin.fauth@uni-tuebingen.de

Prof. Dr. Ilonca Hardy, Faculty of Educational Sciences, Goethe-University, Theodor-W.-Adorno-Platz 6, 60629 Frankfurt am Main, Germany & Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt am Main, Germany
e-mail: hardy@em.uni-frankfurt.de

Prof. Dr. Silke Hertel, Faculty of Behavioural and Cultural Studies, University of Heidelberg, Voßstraße 2, 69115 Heidelberg, Germany
e-mail: hertel@ibw.uni-heidelberg.de

el aggregated student ratings of instructional quality, results show a positive link between classroom social climate and science competence but not for classroom management and science competence. As its most important finding, our study demonstrates the compensatory capacity of instructional quality to narrow the achievement gap between students at risk and their peers. Furthermore, classroom management also counteracted risk of school failure when controlling for students' language proficiency.

Keywords

Instructional quality; Child-by-instruction interaction; Science competence; Children at risk

Zur Bedeutung von Unterrichtsqualität für die naturwissenschaftliche Kompetenz von Grundschulkindern

Ein Fokus auf Kinder mit ungünstigen Lernvoraussetzungen

Zusammenfassung

Im Rahmen von Forschung zu gutem Unterricht wurden Unterrichtsqualitätsmerkmale identifiziert, die mit dem Lernerfolg von Schülerinnen und Schülern verbunden sind. Diese beinhalten bspw. strukturell-organisatorische Aspekte, wie Classroom Management, oder affektive Aspekte, wie das Klassenklima. Dass der Lernerfolg nicht nur von Unterrichtsmerkmalen, sondern auch von deren Wechselspiel mit den individuellen Lernvoraussetzungen abhängt, ist Thema der Forschung zu Aptitude-Treatment-Interactions (ATI). Gegenwärtig erfährt diese Forschung erneute Aufmerksamkeit und bestätigende empirische Befunde. Diese Studie an deutschen Grundschulen (1041 Kinder aus 54 Klassen) knüpft daran an und prüft Haupt- und Interaktionseffekte von Unterrichtsqualitätsmerkmalen (Classroom Management und Klassenklima) sowie individuellen Lernvoraussetzungen (Migrationshintergrund oder geringe kognitive Grundfähigkeiten) für die naturwissenschaftliche Kompetenz von Grundschulkindern. Ergebnisse aus Mehrebenenanalysen zeigen einen positiven Zusammenhang zwischen Klassenklima und naturwissenschaftlicher Kompetenz, nicht aber zwischen Classroom Management und naturwissenschaftlicher Kompetenz. Darüber hinaus zeigten sich Interaktionseffekte zwischen Unterrichtsqualitäts- und Schülermerkmalen dahingehend, dass Unterrichtsqualität einen ausgleichenden Effekt auf die Leistungen von Kindern mit ungünstigen Lernvoraussetzungen und diejenigen ihrer Mitschülerinnen und Mitschüler ausübte. Dieser kompensatorische Effekt zeigte sich für Classroom Management auch nach Kontrolle sprachlicher Kompetenzen.

Schlagworte

Unterrichtsqualität; Aptitude-Treatment-Interaction; Naturwissenschaftliche Kompetenz

1. Introduction

Teachers' classroom instruction is a key to student achievement (Hattie, 2009; Scheerens & Bosker, 1997). Furthermore, research findings have supported the hypothesis that indices of both classroom instruction and individual learner characteristics predict achievement outcomes. Regarding the learners' characteristics, a special focus has been set on students at risk of school failure but the compensatory capacity of instructional quality to counteract risk of school failure needs to be further examined.

2. The relevance of classroom management and classroom social climate for students' achievement

Research on classroom instruction has consistently identified basic dimensions of instructional quality that contribute to student learning (e.g., Klieme, Pauli, & Reusser, 2009; Pianta & Hamre, 2009). These basic dimensions cover, for instance, structural-organizational features such as classroom management, and affective features such as classroom social climate. For several decades, research on instructional quality has acknowledged *classroom management* to be a central feature of successful instruction (e.g., Doyle, 1986; Kounin, 1970). Classroom management includes the implementation of clear rules and procedures in classrooms, effective coping with disruption, and smooth transitions between different instructions and tasks. In effectively managed classrooms, teachers maintain a whole group focus, establish and maintain order, design effective instruction, respond to the needs of individual students, and effectively handle discipline problems (Emmer & Stough, 2001). Classroom management features can be seen as preconditions for students' time-on-task learning, particularly when classroom instruction is characterized by little disruption and smooth transitions, and thus offers more effective learning time (e.g., Seidel & Shavelson, 2007; Wang, Haertel, & Walberg, 1993). After several decades of research on classroom management, there is broad evidence that effective classroom management supports learning of various groups of students and in various domains (e.g., Kunter, Baumert, & Köller, 2007; Wang et al., 1993).

Affective features of instructional quality have been linked to, for instance, the *classroom social climate* that refers to the quality of social relations in classrooms (Mainhard, Brekelmans, & Wubbels, 2011). Classroom social climate covers the enjoyment of being in class and mutual respect displayed during interactions in classrooms. A positive classroom social climate among peers within classes is present when students support and help each other. One of the theoretical approaches to explaining the effect of this social relatedness on achievement is the self-determination theory (Deci & Ryan, 1985; Niemiec & Ryan, 2009). Here, students' intrinsic motivation and engagement, and in turn academic achievement, can be promot-

ed by fulfilling so-called basic needs (i.e., feelings of autonomy, competence, and relatedness). Empirical findings demonstrate that students who feel related and secure in their classes and who have positive peer relationships are more interested in academic activities and are more actively engaged in classroom processes (e.g., Wentzel & Watkins, 2002), which in turn results in higher academic achievement (DeRosier, Kupersmith, & Patterson, 1994; Wentzel, 1993).

3. Science competence and its relation to language proficiency

The development of scientific literacy is an agreed goal of science instruction (e.g., DeBoer, 2000; Smith, Loughran, Berry, & Dimitrakopoulos, 2012). There are several views on scientific literacy such as (a) the understanding of science concepts and their applications, (b) the ability to use scientific knowledge in problem solving, and (c) the knowledge and motivational orientations needed for intelligent participation in science-based social issues (see Bybee, 1997; Norris & Phillips, 2003). In terms of the science education goals, teachers aim to promote an understanding of central scientific concepts, methods of scientific inquiry (including respective epistemological views on the nature of science), an understanding of science in a broader societal context as well as promoting students' motivational and affective resources (e.g., Möller, Hardy, & Lange, 2012). In elementary school, emphasis has been put on the development of scientific knowledge as a transformation of conceptions of core content and principles on a mostly qualitative level of understanding. Thus, elementary science education aims to support students in developing basic and adaptable conceptions in the domains of optics, magnetism, electricity, air, sound, buoyancy, species, and the development of living organisms, among others. This type of science competence does not involve the mere acquisition of isolated facts, but rather the development of a conceptual framework helping students to recognize and interpret patterns and regularities in the natural world.

Science competence is closely connected to language proficiency (e.g., Bos, Wendt, Köller, & Selter, 2012). Science education involves analyzing, summarizing, and presenting information in oral or written formats (Lee & Fradd, 1998). Consequently, language proficiency plays a prominent role in the acquisition of science knowledge in classroom settings. Even in elementary school, science inquiry activities place high demands on students' language proficiency and science vocabulary, because they involve the use of academic discourse in social situations (e.g., when making sense of other students' utterances or when formulating one's own ideas). Consequently, scientific reasoning may pose particular challenges to students with poor language proficiency. Students' development and uses of scientific arguments in classroom discourse has received continuous research interest (e.g., Duschl, 2008; Osborne, Erduran, & Simon, 2004) revealing the intricate relation

between students' scientific reasoning and their uses of language in discourse analysis of classroom interactions or while working on scientific tasks (Lee, 2005).

4. Students at risk of failing to meet the curricular learning goals

Even considering the significance of science learning aims, there is ample evidence that disadvantaged students from minority groups (e.g., from families with immigrant background or lower socio-economic status) show less proficient scientific knowledge than their peers (e.g., Campbell, Hombo, & Mazzeo, 2000), and that this holds especially true for German elementary school children (Bos, Wendt, Köller et al., 2012). In large-scale assessments, students are considered to be at risk when their scores do not exceed the first proficiency level of a test because these students are at risk of failing to complete vocational education (e.g., Maaz & Baumert, 2012). Regarding the conceptualizations of students at risk, two central categories can be distinguished: demographic risks and functional risks of school failure (see Hamre & Pianta, 2005). In terms of demographic risks, research identifies socio-economic status and the particular role of students from *immigrant families*. In Germany, large-scale assessments have revealed that especially students from immigrant families are at risk of failure to reach higher proficiency levels and to complete a vocational education (e.g., Baumert & Maaz, 2012). When focusing on young children, the native status of the parents may be more important than the students' own status because parenting practices and the home environment are important factors affecting the students' early development (De Feyter & Winsler, 2009). Disadvantages of students from immigrant families are often explained by their different cultural background, lower socio-economic status, and in particular by language difficulties (e.g., Bos, Wendt, Köller et al., 2012; Fuligni, 1997; Lee, 2003). Using profile analysis, Bos, Wendt, Ünlü et al. (2012) showed that significantly more students from immigrant families were assigned to the two lowest of seven achievement profiles, and that regarding these lowest achievement profiles language proficiency was particularly poor.

Among the indicators reflecting children's functional risks (e.g., behavioral, social, attentional, and academic problems), *cognitive abilities* are among the most frequently investigated individual determinants of school success or failure. Carroll (1993) defines cognitive abilities as "any ability that concerns some [...] class of tasks in which correct or appropriate processing of mental information is critical to successful performance" (p. 10). Correlation patterns and factor analyses show that general cognitive abilities are strongly connected to specific abilities in various domains at school, especially to mathematics, science and language competencies (e.g., Deary, Strand, Smith, & Fernandes, 2007; Gustafsson & Balke, 1993). For science education, there is ample evidence that the understanding of science concepts and terms is quite complex for all students, and that conceptual learning is a slow

and gradual process (e.g., Caravita & Hallden, 1994). This leads to the assumption that especially students with poor cognitive abilities are at risk of failing to meet the curricular learning goals in science.

5. The importance of instructional quality for students at risk

Based on the theoretical view that students' learning needs may interact with the quality of school learning environments (e.g., Morrison & Connor, 2002), and as intensively debated and investigated within the scope of aptitude-treatment-interactions (ATI) research, the effect of instruction on learning outcomes may differentially depend on student characteristics. In their early review on ATI research, Cronbach and Snow (1977) concluded that there was only suggestive evidence of ATI, and the authors indicated theoretical, educational, and methodological shortcomings of this research tradition. Owing to methodological progress made in the examination of classroom processes such as the use of hierarchical linear modeling (e.g., Raudenbush & Bryk, 2002), attention on child-by-instruction interactions has increased in recent years, as well as respective evidence (e.g., Hamre & Pianta, 2005; Jones & Byrnes, 2006).

5.1 The importance of classroom management for students at risk

Effective classroom management might compensate for at-risk students' disadvantages by providing clear rules and maintaining a scarcely disruptive learning environment (e.g., Finn & Rock, 1997) as well as by structuring of the tasks and the learning content and clarifying the learning goals. Students with functional risks (e.g., poor cognitive abilities or little prior knowledge) are supposed to benefit from more time on task and structured learning settings. Likewise, students with demographic risks (e.g., from immigrant families) are supposed to be supported in structured classrooms with clear rules and few disturbances because these students are considered to feel more confident regarding participation in discourse and learning-related interactions (Curran, 2003). Möller, Jonen, Hardy, and Stern (2002) confirmed that high instructional support (i.e., the sequencing of instructional content and the frequency of cognitively structuring statements) helped students to understand science concepts, and especially lower-achieving students benefited from such instructional support. Building on observations and external ratings of the quality of the environment, Curby, Rimm-Kaufman, and Ponitz (2009) reported that good classroom management helped students with lower levels of mathematics achievement. Cadima, Leal, and Burchinal (2010) confirmed a significant

interaction effect between prior achievement and classroom management on elementary school children's number identification skills.

5.2 The importance of classroom social climate for students at risk

It is well supported that at-risk students participate less fully in learning-related activities and are less fully engaged in learning (e.g., Finn, Pannozzo, & Voelkl, 1995; Finn & Rock, 1997). A positive classroom social climate aims at supporting students' intrinsic motivation and engagement, and in turn academic achievement. Positive classroom interactions and supportive relationships offer ways to academically engage students (e.g., Wentzel & Watkins, 2002) and to increase students' motivation (e.g., Wentzel, Battle, Russell, & Looney, 2010). Hamre and Pianta (2005) showed that emotional support (including aspects of interpersonal relations and classroom management) positively moderated the achievement of students with functional risks of school failure. Curby et al. (2009) reported that students with lower word reading competencies benefited most from high emotional support. Furthermore, research shows that the lower achievement levels of children with demographic risks (e.g., students from immigrant families) may be due to language difficulties as well as sociocultural and psychological factors (e.g., Fuligni, 1997; Portes, 1999). Accordingly, students with demographic risks and thus with social and emotional needs should benefit academically from an appreciative and supportive social climate within the classroom. However, further methodologically sound research, particularly in elementary school, is scarce.

5.3 Methodological considerations to be taken into account when assessing instructional quality

Different sources of assessments have provided evidence for the importance of instructional quality for student learning. Although basic dimensions of instructional quality have often been investigated within the scope of observational studies and assessed via expert ratings (e.g., Klieme et al., 2009; Pianta & Hamre, 2009), from the beginning researchers have also used student ratings of classroom instruction (e.g., Walberg & Anderson, 1968). Yet, the validity of student ratings of classroom instruction has been controversially discussed (e.g., Greenwald, 1997). Other authors advocated the use of student ratings provided that they have been developed in a carefully and theoretically sound manner (e.g., McKeachie, 1997). The correct choice of analysis level is considered as particularly important. Recently Marsh and colleagues (2012) highlighted the classroom level as the relevant unit for the analysis of instructional processes with the aim of identifying differences among classes. Aggregated student ratings of instructional quality can be reliable and valid indicators even in elementary school (Fauth, Decristan, Rieser, Klieme, & Büttner,

2014). Results of the few contemporary studies in elementary schools are inconsistent concerning the connection between student perceptions of instructional quality and achievement (e.g., Fauth et al., 2014; Goh, Young, & Fraser, 1995; LaRocque & Mvududu, 2008).

6. Aims of the study and hypotheses

Despite the large body of research on instructional quality, there is still a lack of research on the relevance of student perceptions of instructional quality for academic achievement, especially in elementary school. Most of all, it is important to examine the extent to which features of instructional quality can help narrow the gap between students at risk of school failure and their peers. With regard to the theoretical assumptions and empirical results laid out, we assume that

(1) students with demographic risks (i.e., students from immigrant families) or functional risks (i.e., students with poor cognitive abilities) show lower science competence levels than their peers, but that disadvantages of students from immigrant families are mainly explained by students' language proficiency;

(2) classroom management and classroom social climate will be generally positively related to students' science competence;

(3) the gap between students at risk of failure in science competence and their peers will be narrower in high-quality classrooms (i.e., with effective classroom management or a positive classroom social climate), and this compensatory effect of high-quality classrooms will remain after controlling for students' language proficiency.

7. Method

7.1 Participants

This analysis was conducted within the framework of the IGEL-study (*Individual support and adaptive learning environments in primary school*), an intervention in German elementary school on the effectiveness of different teaching approaches (see Decristan et al., 2015; Hardy et al., 2011). Schools and teachers were recruited via telephone and information events, and asked to participate in the study. Participation was voluntary for teachers and students. 1,041 third grade elementary school students (49 % female) with a mean age of $M = 8.8$ years ($SD = 0.50$) took part in assessments at the beginning of the academic school year 2010/2011. The sample included diverse ethnic groups, students from immigrant families mostly also speaking Turkish (21 %), Slavic (19 %), Romanic (18 %), African (12 %) or other (Indo-)Germanic (21 %) languages. All schools were located in central Germany, in both rural and urban (68 % of the classes) areas. The overall participation rate

was high (96 % of all students in the 54 classes). The mean class size was 19 students.

7.2 Instruments

7.2.1 Instructional quality

Students' *perception of instructional quality* was assessed via a questionnaire. Students were asked to answer items on a Likert scale ranging from 1 = *strongly disagree* to 4 = *strongly agree*. The scale *classroom management* (see Fauth et al., 2014) covered six items on the occurrence of disciplinary problems and disruption in the classroom (e.g., "In our science class, none of the students disturb the lesson."). Reliability at the individual level was good (Cronbach's $\alpha = .82$). The scale *classroom social climate* covered six items ($\alpha = .80$) that were adapted from Diel and Nieder (2010) focusing on positive relationships between students in class (e.g., "In our science class, we stick together."). Mean scores for classroom management were $M = 2.81$ ($SD = 0.74$) and for classroom social climate $M = 3.20$ ($SD = 0.68$). Student ratings of instructional quality were aggregated to a mean score for each class to be used as a level-2 variable. Because these ratings were used as class-level variables, we calculated indices of intraclass correlations (ICCs). The ICC_1 indicates the proportion of item variance in a sample that can be attributed to differences between groups (here: classes). The ICC_2 describes the reliability of these aggregated scores (e.g., Lüdtke, Trautwein, Kunter, & Baumert, 2006). Results indicate that the scales showed substantial variance between classes and high agreement within classes (classroom management: $ICC_1 = .20$, $ICC_2 = .82$; classroom social climate: $ICC_1 = .25$, $ICC_2 = .86$).

7.2.2 Student variables

The test of students' *science competence* covered seven items adapted from TIMSS 2007 (Martin, Mullis, & Foy, 2008) and five self-developed items. Experts from educational practice and research in science education had judged the items as highly relevant for elementary school science education and appropriate for third-grade students. All items were piloted in class-wide assessments at the end of second grade. The items included the domains of physics, chemistry, and geography, covering items on wind force, seasons, lengths, and aggregation states of water, as well as concepts of density of objects, buoyancy and displacement. All items aimed at testing students' science competence in the respective domains, requiring an application or recognition of basic science principles in contexts of everyday life. The test was scaled to fit the Rasch model (Rasch, 1961). EAP/PV reliability was satisfactory ($r = .70$). WLE-parameters (weighted likelihood estimates; Warm, 1989) were computed for each student. Students' *cognitive abilities* as a function-

al risk indicator were assessed via the CFT 20-R (Weiß, 2006), proven to be a reliable and valid instrument according to a nationwide standardization study. In its short version, it consists of 56 items total (Cronbach's $\alpha = .72$). Students from *immigrant families* as a demographic risk factor were identified via a student questionnaire. Students who reported that either one or both parents were not born in Germany were coded as students from immigrant families (= 1). This was the case for 38 % of the participating students. The other students were coded as not having an immigrant background (= 0). *Language proficiency* was assessed using an instrument and SET 5-10 adapted from diagnostic tests of German language comprehension (Elben & Lohaus, 2000; Glück, 2011; Petermann, Metz, & Fröhlich, 2010). The test of language proficiency consisted of 20 items ($\alpha = .72$) covering both word and sentence comprehension.

7.3 Procedure

In class-wide assessments, data was collected by trained research staff using standardized instructions. Questionnaires and tests were part of two larger surveys lasting about 90 minutes and were conducted on two separate days with $M = 3.5$ days ($SD = 5.1$) in-between, and with instructional quality and science competence being assessed during the first of the two surveys. Instructions and items were read aloud and after each item students were given time to respond. Test items were also visually presented with a projector.

7.4 Data analyses

All analyses were conducted with the software Mplus 7 (Muthén & Muthén, 1998–2012; see Appendix for Mplus commands). Regarding our hypothesis 1, we used stepwise regression analysis with a correction of standard errors to account for the nested data structure. All metric variables were standardized ($M = 0$, $SD = 1$) at the individual level. Furthermore, because we were interested in both students' individual-level variables and instructional quality as classroom level variable, we used multilevel regression analysis (see Appendix for regression equations). Regarding hypothesis 2, we specified a means-as-outcomes model. Scores of instructional quality were standardized ($M = 0$, $SD = 1$) at the classroom level and centered at the grand mean. To test hypothesis 3, we added students' risk indicators. Risk factors that were used to form cross-level interaction terms were centered at the group mean to disentangle the effects between both levels of analysis (see Enders & Tofghi, 2007). Additional individual-level covariates were centered at the grand mean.

In Mplus 7, cases with missing data for any of the manifest predictor variables are not included in the multilevel regression analysis. However, the amount of missing data at the individual level was rather small (ranging between 4 % for

language proficiency and 13 % for students from immigrant families), and there were no missing data at the classroom level (instructional quality). Missing data at the individual level was due to students who were absent during one of the assessments or missed some parts of the assessments.

8. Results

8.1 Hypothesis 1

First, we examined the relevance of students' demographic and functional risk factors for science competence. Correlations between students' immigrant background and language proficiency ($r = -.45$), and cognitive ability and language proficiency ($r = .39$) were substantial.

Table 1: Regression analysis predicting science competence from students' risk factors

	Science Competence											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
Constant	0.34	0.05	0.03	0.05	0.01	0.04	0.14	0.04	0.04	0.03	0.17	0.04
Immigrant background	-0.76*	0.06	-	-	-	-	-0.29*	0.05	-	-	-0.31*	0.05
Cognitive ability	-	-	0.38*	0.03	-	-	-	-	0.19*	0.02	0.21*	0.03
Language proficiency	-	-	-	-	0.58*	0.04	0.51*	0.04	0.50*	0.04	0.42*	0.04
<i>R</i> ²	0.143	-	0.151	-	0.336	-	0.354	-	0.368	-	0.394	-

Note. Students from families with immigrant background = 1.

* $p < .05$ one-tailed.

As Table 1 (Model 1) shows, students from immigrant families scored lower in science competence than their peers. Furthermore, students' cognitive ability was significantly connected with science competence (Model 2). The additional relevance of students' language proficiency is shown in Models 3 to 5. The interplay between students' risk factors and science competence remained significant when controlling for language proficiency. Furthermore, each of the risk variables independently contributed to predicting students' science competence (Model 6).

8.2 Hypothesis 2

Next, we explored the relevance of instructional quality for students' science competence. Contrary to our expectations, classroom management was not significantly connected with students' science competence (Table 2, Model 1). However, as predicted by our hypothesis, we observed a positive link between classroom social climate and students' science competence (Table 3, Model 1). The corresponding amounts of variances explained by instructional quality and effect sizes were small (CM: $R^2 = .026, f^2 = 0.027$; CSC: $R^2 = .108, f^2 = 0.121$).

8.3 Hypothesis 3

Finally, we examined the moderating effects of instructional quality on the interplay between students' risk factors and science competence. As shown in Table 2 (Models 2 and 3), classroom management positively moderated the link between students' demographic risk and science competence and negatively moderated the link between students' functional risk and science competence. Accordingly, classroom management bore the potential to counteract students' risks of school failure (see Figure 1A). These cross-level interactions could still be confirmed when controlling for language proficiency (Models 4 and 5) as well as when considering all individual-level covariates (Models 6 and 7).

Figure 1: Interaction effects of students' risk factors and instructional quality (i.e., (A) classroom management, (B) classroom social climate) on students' science competence scores

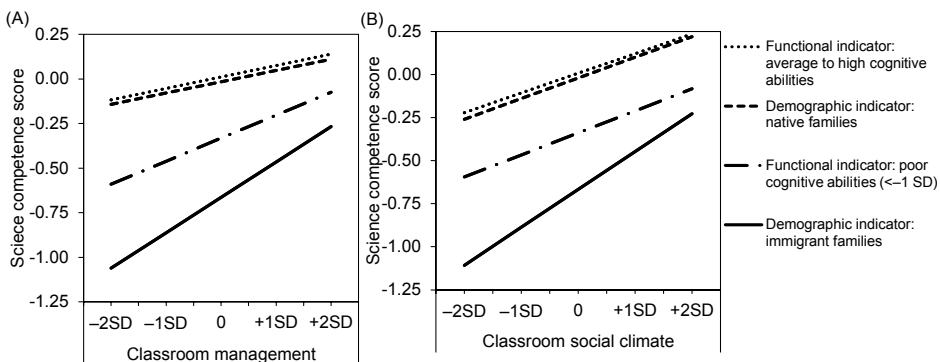


Table 2: Multilevel regression analysis predicting science competence from students' risk factors and classroom management

	Science Competence															
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7			
	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE		
Intercept	-0.02	0.06	-0.02	0.06	0.01	0.06	0.01	0.04	0.03	0.04	0.03	0.04	0.03	0.04	0.03	0.04
<i>Individual-level variables</i>																
Immigrant background	-	-	-0.65*	0.06	-	-	-0.28*	0.05	-	-	-0.29*	0.05	-	-0.30*	0.05	0.05
Cognitive ability	-	-	-	-	0.34*	0.03	-	-	0.19*	0.02	0.21*	0.03	0.03	0.21*	0.03	0.03
Language proficiency	-	-	-	-	-	-	0.51*	0.04	0.51*	0.04	0.43*	0.04	0.43*	0.04	0.43*	0.04
<i>Class-level variable</i>																
CM	0.06	0.05	0.06	0.05	0.06	0.05	0.02	0.04	0.02	0.03	-0.01	0.03	0.03	0.03	0.03	0.03
<i>Interactions between individual-level variables and the class-level variable</i>																
Immigrant background × CM	-	-	0.14*	0.06	-	-	0.12*	0.05	-	-	0.08*	0.05	-	-	-	-
Cognitive ability × CM	-	-	-	-	-0.07*	0.03	-	-	-0.06*	0.02	-	-	-0.06*	0.03	-	-

Note. Students from families with immigrant background = 1. CM = Classroom Management.

* $p < .05$ one-tailed.

Table 3: Multilevel regression analysis predicting science competence from students' risk factors and classroom social climate

	Science Competence													
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Intercept	-0.01	0.06	-0.01	0.06	0.02	0.06	0.01	0.04	0.03	0.04	0.03	0.04	0.16	0.04
<i>Individual-level variables</i>														
Immigrant background	-	-	-0.64*	0.06	-	-	-0.28*	0.06	-	-	-0.29*	0.05	-0.31*	0.05
Cognitive ability	-	-	-	-	0.35*	0.03	-	-	0.19*	0.02	0.21*	0.03	0.21*	0.03
Language proficiency	-	-	-	-	-	-	0.51*	0.04	0.51*	0.04	0.43*	0.04	0.43*	0.04
<i>Class-level variable</i>														
CSC	0.12*	0.06	0.12*	0.06	0.11*	0.05	0.03	0.03	0.03	0.03	-0.01	0.03	0.03	0.03
<i>Interactions between individual-level variables and the class-level variable</i>														
Immigrant background × CSC	-	-	0.10*	0.06	-	-	0.06	0.05	-	-	0.05	0.05	-	-
Cognitive ability × CSC	-	-	-	-	-0.01	0.04	-	-	-0.01	0.03	-	-	-0.01	0.03

Note. Students from families with immigrant background = 1. CSC = Classroom Social Climate.

* $p < .05$ one-tailed.

The moderating effects of classroom social climate on the interplay between students' risk factors and science competence are shown in Table 3 (Models 2 and 3). Classroom social climate positively moderated the negative connection between immigrant background and science competence, but did not moderate the positive relation between cognitive ability and science competence. Consequently, students from immigrant families performed better in classes with a positive classroom social climate (see Figure 1B). However, after controlling for students' language proficiency and further individual variables, this child-by-instruction interaction no longer proved to be significant (Models 4 to 7).

9. Discussion

This paper examined the relevance of instructional quality for students' science competence with a focus on the academic benefit of instructional quality, particularly for students at risk of school failure. Based on elementary school students' perceptions of instructional quality and using multilevel regression analysis, we demonstrated a positive link between classroom social climate and students' science competence. Furthermore and foremost, this paper extends previous research on instructional quality by showing the potential of classroom management to counteract risk of school failure by narrowing the gap in science competence between students at risk and their peers.

Results regarding our hypothesis 1 supported previous findings on the strong connection between students' demographic and functional risks and science competence (e.g., Bos, Wendt, Köller et al., 2012; Campbell et al., 2000). Furthermore, these results point to the relevance of language in science education and the need to design science curricula that enable students to use language by describing, explaining, and reflecting scientific concepts (Hachfeld, Anders, Schroeder, Stanat, & Kunter, 2010). Students with language difficulties may academically benefit from concrete experiences and opportunities to use language in academic and social settings (e.g., Lee & Fradd, 1998) given that these settings offer adequate support and structure for the use of specific linguistic content. The importance of content-embedded language has been emphasized in programs focusing on second language acquisition (e.g., Darsow, Paetsch, Stanat, & Felbrich, 2012). Nevertheless, the interplay between students' risk indicators and science competence remained significant when controlling for language proficiency. This finding supports previous results on the (additional) importance of (socio-)cultural and psychological factors for the achievement of students at risk (e.g., Fuligni, 1997; Portes, 1999) that should be addressed by teachers. Furthermore, as reported in the section "participants", students from different language backgrounds may have had different chances to answer the test items. Consequently, differential item functioning in future analyses may provide further insights into this issue.

Regarding the relevance of instructional quality for students' science competence (hypothesis 2), we only confirmed a positive link between classroom social climate and science competence. As the beneficial impact of effective classroom management on achievement has been described in various other studies (e.g., Emmer & Stough, 2001; Wang et al., 1993), the fact that we did not find a link between classroom management and science competence here may be attributable to our operationalization of classroom management. Our student questionnaire referred to the non-occurrence of disciplinary problems, whereas other studies (e.g., Pianta & Hamre, 2009) also assessed the strategies teachers use to provide structure in the classroom and to convey clear expectations. The mere maintenance of discipline in the classroom might not be sufficient to foster learning. Furthermore, both the missing effect of classroom management and the small effect of classroom social climate on science achievement might be explained by the wording of our items: We asked for the German subject '*Sachunterricht*' (which includes natural science as well as social sciences), whereas our criterion was science competence. The effects would probably be more pronounced when asking more specifically for science education. In a longitudinal analysis, Fauth et al. (2014) confirmed an effect of classroom management on science achievement as related to a specified science unit.

Finally, regarding our interaction hypothesis, we confirmed that effective classroom management in elementary school narrowed the gap in science competence between students from both immigrant families and with poor cognitive abilities and their peers. Students at risk particularly benefited from classroom instruction with relatively few disruptions. These results add to previous research (e.g., Curran, 2003; Möller et al., 2002) in addressing the elementary students' perspective on instructional quality. Furthermore, classroom management also counteracted risk for school failure when considering students' language proficiency. Consequently, students with lower levels of cognitive ability or with immigrant background academically benefited from good classroom management regardless of their language proficiency.

Concerning the expected interactions between classroom social climate and students' risk factors, our hypothesis was not fully confirmed. Students from immigrant families particularly benefited from a positive classroom climate. Contrary to previous findings (e.g., Curby et al., 2009; Hamre & Pianta, 2005), the results of our study could not confirm a child-by-instruction interaction for students with lower levels of cognitive ability. Moreover, when controlling for students' language proficiency both the main and interaction effects of classroom social climate no longer proved to be significant. Consequently, these results also illustrate the interplay between classroom social climate and language proficiency. It might be easier to implement a positive social climate in classes with high language proficiency scores. Alternatively, in classes with a positive social climate students are supposed to feel more confident, and this might also provide opportunities for the use and development of language.

This leads to the limitations of our paper. Because of the cross-sectional data analysis, results can only focus on the interplay of instructional quality, students' risk factors, and science competence. Ratings of instructional quality might be biased by students' individual characteristics. Although previous research has shown that neither gender nor socio-economic status was connected with individuals' perceptions of instructional quality (e.g., Brock, Nishida, Chiong, Grimm, & Rimm-Kaufman, 2008; Doll, Spies, LeClair, Kurien, & Foley, 2010), this bias has not yet been examined for ratings of students from immigrant families or with poorer cognitive abilities. Consequently, when classes widely differ in the proportion of students at risk, there might be a systematic bias of class-level aggregated ratings of instructional quality (e.g., Sanford & Evertson, 1981). In turn, classroom management or classroom social climate might also differ depending on the proportion of children at risk. Therefore, the (reciprocal) effects between instructional quality, students at risk, and science competence need to be examined in future analyses.

Regarding our assessment of instructional quality, the validity of student ratings has been discussed (e.g., Greenwald, 1997). The instruments were carefully developed, and they showed sufficiently reliable results. The validity of the applied classroom management scale has just recently been confirmed (Fauth et al., 2014). Methodological progress in the examination of student ratings (i.e., hierarchical linear modeling and aggregated student ratings) was also taken into account.

In summary, recent methodologically sound studies point to the value of research on child-by-instruction interactions (e.g., Cadima et al., 2010; Curby et al., 2009; Hamre & Pianta, 2005). Our paper adds to this body of ATI-research and emphasizes the compensatory capacity of classroom management to counteract risk for school failure.

Acknowledgements

This research was funded by the Hessian Initiative for the Development of Scientific and Economic Excellence (LOEWE).

References

- Baumert, J., & Maaz, K. (2012). Migration und Bildung in Deutschland [Migration and education in Germany]. *Die Deutsche Schule*, 104(3), 279–302.
- Bos, W., Wendt, H., Köller, O., & Selzer, C. (Eds.). (2012). *TIMSS 2011 – Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* [TIMSS 2011. Mathematics and science competencies of German elementary school students in international comparison]. Münster, Germany: Waxmann.
- Bos, W., Wendt, H., Ünlü, A., Valtin, R., Euen, B., Kasper, D., & Tarelli, I. (2012). Leistungsprofile von Viertklässlerinnen und Viertklässlern in Deutschland [Achievement profiles of fourth graders in Germany]. *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 269–301). Münster, Germany: Waxmann.

- Brock, L. L., Nishida, T. K., Chiong, C., Grimm, K. J., & Rimm-Kaufman, S. E. (2008). Children's perceptions of the classroom environment and social and academic performance: A longitudinal analysis of the contribution of the responsive classroom approach. *Journal of School Psychology, 46*(2), 129–149. doi:10.1016/j.jsp.2007.02.004
- Bybee, R. W. (1997). Toward an understanding of scientific literacy. In W. Gräber, & C. Bolte (Eds.), *Scientific literacy* (pp. 37–68). Kiel: IPN.
- Cadima, J., Leal, T., & Burchinal, M. (2010). The quality of teacher-student interactions: Associations with first graders' academic and behavioral outcomes. *Journal of School Psychology, 48*(6), 457–482. doi:10.1016/j.jsp.2010.09.001
- Campbell, J. R., Hombro, C. M., & Mazzeo, J. (2000). *NAEP 1999 trends in academic progress: Three decades of student performance (NCES 2000-469)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Caravita, S., & Hallden, O. (1994). Re-framing the problem of conceptual change. *Learning and Instruction, 4*(1), 89–111. doi:10.1016/0959-4752(94)90020-5
- Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor-analytic studies*. New York, NY: Cambridge University Press.
- Cronbach, J. L., & Snow, E. R. (1977). *Aptitudes and instructional methods – a handbook for research on interactions*. New York, NY: Irvingston.
- Curby, T. W., Rimm-Kaufman, S. E., & Ponitz, C. C. (2009). Teacher-child interactions and children's achievement trajectories across kindergarten and first grade. *Journal of Educational Psychology, 101*(4), 912–925. doi:10.1037/a0016647
- Curran, M. E. (2003). Linguistic diversity and classroom management. *Theory into practice, 42*, 334–340. doi:10.1353/tip.2003.0042
- Darsow, A., Paetsch, J., Stanat, P., & Felbrich, A. (2012). Ansätze der Zweitsprachförderung: Eine Systematisierung [Approaches to promote second language: a systematization]. *Unterrichtswissenschaft, 40*(1), 64–82.
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13–21. doi:10.1016/j.intell.2006.02.001
- DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching, 37*(6), 582–601. doi:10.1002/1098-2736(200008)37:6<582::AID-TEA5>3.0.CO;2-L
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum.
- Decristan, J., Hondrich, A. L., Büttner, G., Hertel, S., Klieme, E., Kunter, M., Lühken, A., Adl-Amini, K., Djakovic, S.-K., Mannel, S., Naumann, A., & Hardy, I. (2015). Impact of additional guidance in science education on primary students' conceptual understanding. *The Journal of Educational Research, 108*(5), 358–370. doi:10.1080/00220671.2014.899957
- De Feyter, J. J., & Winsler, A. (2009). The early developmental competencies and school readiness of low-income, immigrant children: Influences of generation, race/ethnicity, and national origins. *Early Childhood Research Quarterly, 24*(4), 411–431. doi:10.1016/j.ecresq.2009.07.004
- DeRosier, M. E., Kupersmith, J. B., & Patterson, C. J. (1994). Children's academic and behavioral adjustment as a function of the chronicity and proximity of peer rejection. *Child Development, 65*(6), 1799–1813. doi:10.2307/1131295
- Diel, E., & Nieder, T. (2010). *Fragebögen zum Klassenklima. Hessischer Referenzrahmen Schulqualität (HRS), Qualitätsbereiche V „Schulkultur“ und VI „Lehren und Lernen“* [Questionnaire on classroom social climate]. Wiesbaden, Germany: Institut für Qualitätsentwicklung.
- Doll, B., Spies, R. A., LeClair, C. M., Kurien, S. A., & Foley, B. P. (2010). Student perceptions of classroom learning environments: Development of the ClassMaps survey. *School Psychology Review, 39*(2), 203–218.

- Doyle, W. (1986). Classroom organization and management. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 392–431). New York, NY: Macmillan.
- Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, 32(1), 268–291. doi:10.3102/0091732X07309371
- Elben, C. E., & Lohaus, A. (2000). *Marburger Sprachverständnistest für Kinder ab 5 Jahren* [Marburger test of language comprehension for children 5 years of age and older]. Göttingen, Germany: Hogrefe.
- Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, 36(2), 103–112. doi:0.1207/S15326985EP3602_5
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121–138. doi:10.1037/1082-989X.12.2.121
- Fauth, B. C., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9. doi:10.1016/j.learninstruc.2013.07.001
- Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive-withdrawn behavior and achievement among fourth graders. *The Elementary School Journal*, 95(5), 421–434. doi:10.1086/461853
- Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk of school failure. *Journal of Applied Psychology*, 82(2), 221–234. doi:10.1037/0021-9010.82.2.221
- Fulgini, A. J. (1997). The academic achievement of adolescents from immigrant families: The role of family background, attitudes, and behavior. *Child Development*, 68(2), 351–363. doi:10.2307/1131854
- Glück, C. W. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige: WWT 6–10 [Test of vocabulary of 6- to 10-year-old children]*. Munich, Germany: Elsevier, Urban & Fischer.
- Goh, S. C., Young, D. J., & Fraser, B. J. (1995). Psychosocial climate and student outcomes in elementary mathematics classrooms: A multilevel analysis. *The Journal of Experimental Education*, 64(1), 29–40. doi:10.1080/00220973.1995.9943793
- Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52(11), 1182–1186. doi:10.1037//0003-066X.52.11.1182
- Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28(4), 407–434. doi:10.1207/s15327906mbr2804_2
- Hachfeld, A., Anders, Y., Schroeder, S., Stanat, P., & Kunter, M. (2010). Does immigration background matter? How teachers' predictions of students' performance relate to student background. *International Journal of Educational Research*, 49(2), 78–91. doi:10.1016/j.ijer.2010.09.002
- Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, 76(5), 949–967. doi:10.1111/j.1467-8624.2005.00889.x
- Hardy, I., Hertel, S., Kunter, M., Klieme, E., Warwas, J., Büttner, G., & Lühken, A. (2011). Adaptive Lerngelegenheiten in der Grundschule: Merkmale, methodisch-didaktische Schwerpunktsetzungen und erforderliche Lehrerkompetenzen [Adaptive learning opportunities in elementary school]. *Zeitschrift für Pädagogik*, 57(6), 819–833.
- Hattie, J. A. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

- Jones, K. K., & Byrnes, J. P. (2006). Characteristics of students who benefit from high-quality mathematics instruction. *Contemporary Educational Psychology, 31*(3), 328–343. doi:10.1016/j.cedpsych.2005.10.002
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Münster, Germany: Waxmann.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York, NY: Holt, Rinehart & Winston.
- Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction, 17*(5), 494–509. doi:10.1016/j.learninstruc.2007.09.002
- LaRocque, M., & Mvududu, N. (2008). Assessing perceptions of the environment in elementary classrooms: The link with achievement. *Educational Psychology in Practice, 24*(4), 289–305. doi:10.1080/02667360802488732
- Lee, O. (2003). Equity for linguistically and culturally diverse students in science education: A research agenda. *Teachers College Record, 105*(3), 465–489. doi:10.1111/1467-9620.00247
- Lee, O. (2005). Science education with English language learners: Synthesis and research agenda. *Review of Educational Research, 75*(4), 491–530. doi:10.3102/00346543075004491
- Lee, O., & Fradd, S. H. (1998). Science for all, including students from non-English language backgrounds. *Educational Researcher, 27*(4), 12–21. doi:10.2307/1176619
- Lüdtke, O., Trautwein, U., Kunter, M., & Baumert, J. (2006). Reliability and agreement of student ratings of the classroom environment: A reanalysis of TIMSS data. *Learning Environments Research, 9*(3), 215–230. doi:10.1007/s10984-006-9014-8
- Maaz, K., & Baumert, J. (2012). Risikoschüler in Deutschland [Students' at risk in Germany]. In S. G. Huber (Ed.), *Jahrbuch Schulleitung 2012. Befunde und Impulse zu den Handlungsfeldern des Schulmanagements* (pp. 77–89). Köln, Germany: Carl Link.
- Mainhard, M. T., Brekelmans, M., & Wubbels, T. (2011). Coercive and supportive teacher behavior: Within- and across-lesson associations with the classroom social climate. *Learning and Instruction, 21*(3), 345–354. doi:10.1016/j.learninstruc.2010.03.003
- Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47*(2), 106–124. doi:10.1080/00461520.2012.670488
- Martin, M. O., Mullis, I. V. S., & Foy, P. (with Olson, J. F., Erberber, E., Preuschoff, C., & Galia, J.). (2008). *TIMSS 2007 international science report: findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist, 52*(11), 1218–1225. doi:10.1037/0003-066X.52.11.1218
- Möller, K., Hardy, I., & Lange, K. (2012). Moving beyond standards: How can we improve elementary science learning? A German perspective. In S. Bernholt, K. Neumann, & P. Nentwig (Eds.), *Making it tangible – learning outcomes in science education* (pp. 33–58). Münster, Germany: Waxmann.
- Möller, K., Jonen, A., Hardy, I., & Stern, E. (2002). Die Förderung von naturwissenschaftlichem Verständnis bei Grundschulkindern durch Strukturierung der Lernumgebung [Support of elementary school students' science understanding through the structuring of the learning environment]. In M. Prenzel & J. Doll

- (Eds.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (pp. 176–191). Weinheim, Germany: Beltz.
- Morrison, F. J., & Connor, C. M. (2002). Understanding schooling effects on early literacy: A working research strategy. *Journal of School Psychology, 40*(6), 493–500. doi:10.1016/S0022-4405(02)00127-9
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus (Version 7)* [Computer Software]. Los Angeles, CA: Muthén & Muthén.
- Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education, 7*(2), 133–144. doi:10.1177/1477878509104318
- Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87*(2), 224–240. doi:10.1002/sci.10066
- Osborne, J. F., Erduran, S., & Simon, S. (2004). Enhancing the quality of argument in school science. *Journal of Research in Science Teaching, 41*(10), 994–1020. doi:10.1002/tea.20035
- Petermann, F., Metz, D., & Fröhlich, L. P. (2010). *SET 5–10. Sprachstandserhebungstest für Kinder im Alter zwischen 5 und 10 Jahren* [Test of language comprehension of 5- to 10-year-old children]. Göttingen, Germany: Hogrefe.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119. doi:10.3102/0013189X09332374
- Portes, P. (1999). Social and psychological factors in the academic achievement of children of immigrants: A cultural history puzzle. *American Educational Research Journal, 36*(3), 489–507. doi:10.3102/00028312036003489
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, Volume 4: Contributions to biology and problems of medicine* (pp. 321–333). Berkeley, CA: University of California Press.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.
- Sanford, J. P., & Evertson, C. M. (1981). Classroom management in a low SES junior high: Three case studies. *Journal of Teacher Education, 32*(1), 34–38. doi:10.1177/002248718103200108
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford, United Kingdom: Pergamon.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis research. *Review of Educational Research, 77*(4), 454–499. doi:10.3102/0034654307310317
- Smith, K. V., Loughran, J., Berry, A., & Dimitrakopoulos, C. (2012). Developing scientific literacy in primary school. *International Journal of Science Education, 34*(1), 127–152. doi:10.1080/09500693.2011.565088
- Walberg, H. J., & Anderson, G. J. (1968). Classroom climate and individual learning. *Journal of Educational Psychology, 59*(6), 414–419. doi:10.1037/h0026490
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*(3), 249–294. doi:10.3102/00346543063003249
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450. doi:10.1007/BF02294627
- Weiß, R. H. (2006). *CFT 20-R. Grundintelligenztest Skala 2 – Revision* [Culture Fair Test]. Göttingen, Germany: Hogrefe.
- Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology, 85*(2), 357–364. doi:10.1037//0022-0663.85.2.357

- Wentzel, K. R., Battle, A., Russell, S. L., & Looney, L. B. (2010). Social supports from teachers and peers as predictors of academic and social motivation. *Contemporary Educational Psychology, 35*(3), 193–202. doi:10.1016/j.cedpsych.2010.03.002
- Wentzel, K. R., & Watkins, D. E. (2002). Peer relationships and collaborative learning as contexts for academic enablers. *School Psychology Review, 31*(3), 366–377.

Appendix

Regression equations

Example for hypothesis 1

$$\text{Level 1: } y_i = \beta_0 + \beta_1 \times MB_i + \beta_2 \times (LP_i - \overline{LP}) + \varepsilon_i \quad (1)$$

The outcome y for each person i is predicted from the regression constant β_0 and the regression coefficients β_1 and β_2 of both independent variables MB (immigration background) and LP (language proficiency), and an error term ε . The variable LP is centered at the grand mean (i.e., the mean score of the sample is subtracted from each individual score).

Example for hypothesis 2

$$\text{Level 1: } y_{ij} = \beta_{0j} + r_{ij} \quad (2.1)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01} \times (\overline{CM}_j - \overline{CM}) + u_{0j} \quad (2.2)$$

The outcome y for each person i in each class j is predicted from the intercept (γ_{00}) and the slope (γ_{01}) of the independent variable CM (classroom management), and individual-level and class-level error terms r and u . CM is centered at the grand mean (i.e., the average of the class means is subtracted from each class mean).

Example for hypothesis 3

$$\text{Level 1: } \hat{y}_{ij} = \beta_{0j} + \beta_{1j} \times (MB_{ij} - \overline{MB}_j) + \beta_{2j} \times (LP_i - \overline{LP}) + r_{ij} \quad (3.1)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01} \times (\overline{CM}_j - \overline{CM}) + u_{0j} \quad (3.2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} \times (\overline{CM}_j - \overline{CM}) + u_{1j}$$

$$\beta_{2j} = \gamma_{20}$$

The outcome y for each person i in each class j is predicted from intercepts (γ_{00} , γ_{10} , γ_{20}), the slope (γ_{01}) of the independent variable CM , a cross-level interaction term β_{1j} (i.e., the interaction of the individual-level variable MB and the class level variable CM), and individual-level and class-level error terms r and u . At the classroom level, CM is centered at the grand mean. At the individual level, MB is centered at the group mean (i.e., the mean score of the class is subtracted from each individual score belonging to that class) and LP is centered at the grand mean.

Mplus 7 commands

Example for hypothesis 1

```
Variable:      names = class sc mb ca lp cm_c csc_c;
              usevar = sc mb ca lp;
              cluster = class;
              missings = all (999);
Define:       center ca lp (grandmean);
Analysis:     type = complex;
Model:        sc on mb ca lp;
```

Example for hypothesis 2

```
Variable:      names = class sc mb ca lp cm_c csc_c;
              usevar = sc cm_c;
              between = cm_c;
              cluster = class;
              missings = all (999);
Define:       center cm_c (grandmean);
Analysis:     type = twolevel;
Model:        %between%
              sc on cm_c;
```

Example for hypothesis 3

```
Variable:      names = class sc mb ca lp cm_c csc_c;
              usevar = sc mb lp cm_c;
              within = mb lp;
              between = cm_c;
              cluster = class;
              missings = all (999);
Define:       center mb (groupmean);
              center lp cm_c (grandmean);
Analysis:     type = twolevel random;
Model:        %within%
              betaij | sc on mb;
              sc on lp;
              %between%
              sc on cm_c;
              betaij on cm_c;
```